

28. The method of claim 27, wherein access to each of the plurality of host servers is controlled by one of a plurality of computers, and

wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers.

29. The method of claim 28, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.

30. The method of claim 27, wherein the receiving comprises receiving a request from the client to resolve a domain name associated with any one of the plurality of host servers.

31. The method of claim 27, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of one or more of the plurality of host servers.

32. The method of claim 27, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of an application on one or more of the plurality of host servers.

33. The method of claim 28, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.

34. The method of claim 27, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including a geographical location associated with one or more of the plurality of host servers.

35. The method of claim 28, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of the available session capacity of the respective computer.

36. The method of claim 28, wherein the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by the respective computer to provide an indication of the health of a host server access to which is controlled by the respective computer.

37. The method of claim 28, wherein the identifying comprises identifying one or more of the plurality of host computers based on performance metrics collected from each

of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by the respective computer to provide an indication of the health of an application on a host server access to which is controlled by the respective computer.

38. The method of claim 27, wherein the identifying comprises identifying one or more of the plurality of host servers based on the number of times each of the one or more plurality of host servers has been previously identified.

39. The method of claim 27, wherein the sending comprises sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers.

40. The method of claim 39, wherein the sending comprises sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers, and wherein the one or more network addresses are ordered based on the performance metrics.

41. A system for providing load balancing among a plurality of host servers in a computer network, the system comprising:

a plurality of computers each of which controls access to one or more of the plurality of host servers;

wherein each of the plurality of computers collects performance metrics; and

wherein at least one of the plurality of computers is configured to receive the performance metrics collected by each of the other computers of the plurality of computers and, in response to a request received from a client relating to any one of the plurality of host servers, to identify one or more of the plurality of host servers based on the performance metrics collected by itself and received from each of the other computers of the plurality of computers and to send a response to the client including information associated with each of the one or more identified host servers.

42. A system for providing load balancing among a plurality of host servers in a computer network, the system comprising:

means for receiving a request from a client relating to any one of the plurality of host servers;

means for identifying one or more of the plurality of host servers based on performance metrics including at least a round trip time associated with the client; and

means for sending a response to the client including information associated with each of the one or more identified host servers.

43. The system of claim 42, wherein access to each of the plurality of host servers is controlled by one of a plurality of computers, and

wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers.

44. The system of claim 43, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.

45. The system of claim 42, wherein the means for receiving comprises means for receiving a request from the client to resolve a domain name associated with any one of the plurality of host servers.

46. The system of claim 42, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of one or more of the plurality of host servers.

47. The system of claim 42, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of an application on one or more of the plurality of host servers.

48. The system of claim 43, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.

49. The system of claim 42, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics including a geographical location associated with one or more of the plurality of host servers.

AI  
50. The system of claim 43, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of the available session capacity of the respective computer.

51. The system of claim 43, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by the respective computer to provide an indication of the health of a host server access to which is controlled by the respective computer.

52. The system of claim 43, wherein the means for identifying comprises means for identifying one or more of the plurality of host computers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by the respective computer to provide an indication of the health of an application on a host server access to which is controlled by the respective computer.

53. The system of claim 42, wherein the means for identifying comprises means for identifying one or more of the plurality of host servers based on the number of times each of the one or more plurality of host servers has been previously identified.

54. The system of claim 42, wherein the means for sending comprises means for sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers.

55. The system of claim 54, wherein the means for sending comprises means for sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers, and wherein the one or more network addresses are ordered based on the performance metrics.

56. A computer program product comprising a computer readable medium having computer readable code embodied therein, the computer readable code, when executed, causing a computer to implement a method for providing load balancing among a plurality of host servers in a computer network, the method comprising:

receiving a request from a client relating to any one of the plurality of host servers;

identifying one or more of the plurality of host servers based on performance metrics including at least a round trip time associated with the client; and

sending a response to the client including information associated with each of the one or more identified host servers.

57. The computer program product of claim 56, wherein access to each of the plurality of host servers is controlled by one of a plurality of computers, and

wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers.

58. The computer program product of claim 57, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.



59. The computer program product of claim 56, wherein, in the implemented method, the receiving comprises receiving a request from the client to resolve a domain name associated with any one of the plurality of host servers.

60. The computer program product of claim 56, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of one or more of the plurality of host servers.

61. The computer program product of claim 56, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including an indication of the health of an application on one or more of the plurality of host servers.

62. The computer program product of claim 57, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of whether the number of sessions communicating through the respective computer exceeds a predetermined threshold.

63. The computer program product of claim 56, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics including a geographical location associated with one or more of the plurality of host servers.

64. The computer program product of claim 57, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include an indication of the available session capacity of the respective computer.

65. The computer program product of claim 57, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by the respective computer to provide an indication of the health of a host server access to which is controlled by the respective computer.

66. The computer program product of claim 57, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host computers based on performance metrics collected from each of the plurality of computers, wherein the performance metrics collected from each of the plurality of computers include a time required by

the respective computer to provide an indication of the health of an application on a host server access to which is controlled by the respective computer.

67. The computer program product of claim 56, wherein, in the implemented method, the identifying comprises identifying one or more of the plurality of host servers based on the number of times each of the one or more plurality of host servers has been previously identified.

68. The computer program product of claim 56, wherein, in the implemented method, the sending comprises sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers.

69. The computer program product of claim 68, wherein, in the implemented method, the sending comprises sending a response to the client including information associated with each of the one or more identified servers, wherein the information includes one or more network address each of which is associated with one of the one or more identified host servers, and wherein the one or more network addresses are ordered based on the performance metrics.